Middle-East Journal of Scientific Research 23 (8): 1626-1631, 2015
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2015.23.08.22413

Gravitional Search Optimization for the User Authentication in Biometrics

[1]V. Chandrasekar, [2]M. Akila and [3]T. Maheswari

[1]Department of Computer Science and Engineering,
Vivekanandha College of Technology for Women, Namakkal, India
[2]Department of Information Technology,
[3]Department of Computer Science and Engineering,
Vivekanandha College of Engineering for Women, Namakkal, India

Abstract: The unauthorized users are discriminated using an effective technique called keystroke dynamics which is the analysis of typing rhythm of the users. To develop a repeatable evaluation procedure a number of keystroke dataset is to be collected which used to detect the impostors. We collected a sample dataset from 52 users typing the same passwords each for 10 times and stored in the template using the preprocessing techniques called Hausdorff timing. The gravitational search optimization is used as a feature selection and it is used to compare the preprocessing technique Hausdorff timing with mean, median and standard deviation.

Key words: Biometrics · Keystroke dynamics · Preprocessing · Feature selection · Classification · Hausdorff timing · Gravitational search optimization

INTRODUCTION

The possession of login ID and password are used by most computer systems to prevent user access (i.e. impostors). If the detail is exposed to the impostor then they can access the computer system which will result in security breaches [1-20]. So the transactions based on computers through internet become vulnerable. So the computer security enhanced to provide biometric security. The attributes based on physiological and behavioral are the two major forms of biometrics. Physiological biometrics is biological/chemical traits that are innate or naturally grown for example face, palm, iris recognition, etc. and behavioral biometrics are mannerisms or traits that are learned or acquired for example voice, handwriting, etc. In biometrics there are two phases namely enrollment phase and verification phase [20, 3].

As shown in Figure 1, the features are processed and stored in database during enrollment phase which act as template for further purpose. During the verification phase the same features are extracted and they are compared with the stored template and if it is matched then the user is authorized otherwise unauthorized. The complex pattern recognition do not involve in any other conventional methods [21-23]. There are two types of error metrics namely false positive and false negative. A false positive refers to identifying impostor to be a genuine user. A false negative refers to the rejection of a genuine user considering the user as an impostor. To achieve a high overall accuracy of the system it is desirable to maintain both errors as low.

Keystroke Dynamics: Keystroke dynamics is a behavioral measurement which aims to identify user based on their typing rhythm. The principle behind keystroke dynamics is to extract and analyze the way an individual types as opposed to only what the individual types [17]. The typing rhythm includes how long a key is pressed or released. The duration, latency, digraph, Trigraph and n-graph are the attributes used to measure the keystroke values of a user as shown in Figure 2 [24-28].

Some of the advantages of keystroke dynamics are

- Cheaper to implement
- More distributed
- More unobtrusive
- Requires simply a keyboard and basic software

Corresponding Author: V. Chandrasekar, Department of Computer Science and Engineering, Vivekanandha College of Technology for Women, Namakkal, India. E-mail: drchandru86@gmail.com.

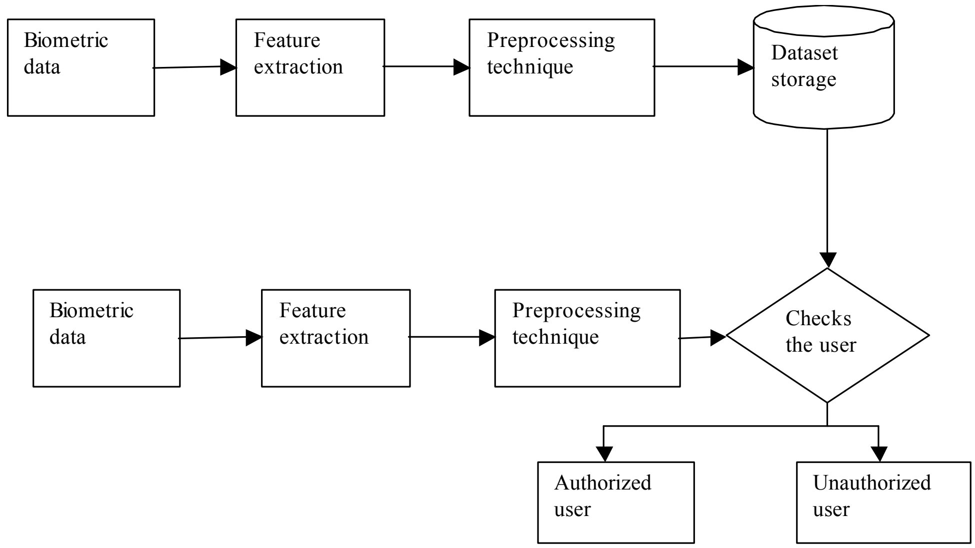

Fig. 1: Phases of biometrics system

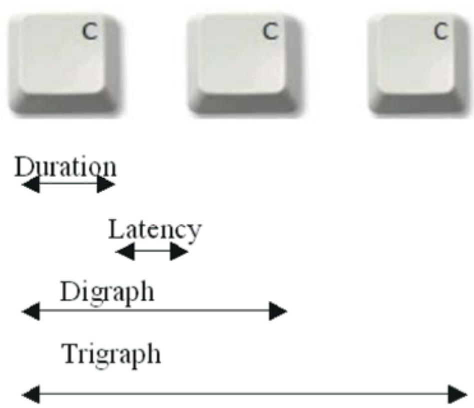

Fig. 2: Keystroke timing information

Some of the disadvantages of keystroke dynamics are

- Typing rhythm are inconsistent
- Typing patterns vary based on the type of keyboard being used.

Feature Extraction: Feature collection is the process of collecting the user's data and storing in the template. The sensor unit senses the keyboard timing of the user. The preprocessing processes some of the data which are sensed by the sensor unit. The feature exactor is the process whereby unique data are extracted from the sample and a template is created. The template for any two persons should differ whereas different samples for the same person should be identical. The preprocessing techniques used here Hausdorff timing, Mean, median and Standard Deviation. The Hausdorff timing is given by [1, 12, 15, 16, 21, 27].

$$H = \left|\sqrt{\left(\sum_{\max i=1}^{n} \sum_{\min j=1}^{n} (p_i - q_j)^2\right)}\right| \quad (1)$$

The other preprocessing techniques are the mean (μ_i), median (m) and standard deviation (σ_i) which are given using the equation [13],

$$\mu_i = \frac{1}{N} \sum_{j=1}^{N} f_j \quad (2)$$

$$m = \left(\frac{N+1}{2}\text{value}\right) \text{when N is odd} \quad (3)$$

$$m = \text{average of} \left(\frac{N}{2}\text{value and } \frac{N}{2}+1\right) \text{when N is Even} \quad (4)$$

$$\sigma_i = \sqrt{\left(\frac{N}{2} \sum_{i,j=1}^{N} |f_j - \mu_i|\right)} \quad (5)$$

Feature Selection: Feature selection is an important technique after preprocessing for effective data analysis in many areas especially classification. Relevant features are usually different to determine without prior knowledge in classification. So the feature selection used in this paper is Gravitational Search algorithm.

Gravitational search algorithm (GSA) is developed based on the law of gravity and mass interactions. According to Newton law, each particle in the universe attracts other particle. In GSA a population of solutions (agents) is maintained. In GSA, agents are considered as objects and their masses are used for measuring their performance. All objects move towards the objects with heavier masses. Heavy masses move more slowly and are known as good solutions. There are three kinds of masses which are taken as best solutions. M_a is active gravitational mass which shows the strength of the gravitational field because of a particular object. M_p Is passive gravitational mass which is related to the strength of an object's interaction with the gravitational field? Inertial mass or M_i, is related to the resistance of an object to changing its state of motion when a force is applied. The agents with bigger inertia mass have slower motion in search space and a more accurate search while the bigger gravitational mass has faster convergence due to having a higher attraction [6, 11, 18, 22, 25]. Each mass (agent) in GSA also has position corresponding to the solution of a problem. Also, a fitness function is used to determine the gravitational and inertial masses. In equation (6), the position of i^{th} agent in the d^{th} dimension is i^d.

$$X_i=(x_i^1,...,x_i^d,...,x_i^n) \text{ For i=1, 2, N} \tag{6}$$

In equation (7), F_{ij}^d (t) is the force on mass 'i' from mass 'j' at time't'. M_{aj} Is the active gravitational mass of agent j? The passive gravitational mass of agent i am M_{pi}. Gravitational constant at time t is G (t). □ is a small constant and R_{ij}(t) shows the Euclidian distance between agents i and j. It is defined as in equation (8).

$$F_{ij}^d(t) = G(t)\frac{M_{pi}(t)\times M_{aj}(t)}{R_{ij}(t)+s}\left(x_j^d(t)-x_i^d(t)\right) \tag{7}$$

$$R_{ij}(t) = \| X_i(t), X_j(t)\|_2 \tag{8}$$

$rand_j$ Shows a random number in the interval [0, 1]. The total force on agent i is as in equation (9).

$$F_i^d(t)\sum\nolimits_{j=1, j\neq i}^{N} rand_j F_{ij}^d(t) \tag{9}$$

The fitness function is calculated for each solutions obtained and the best, worse of the maximum and minimum values are calculated. The function of GSA is as follows: 1) Initialize population, 2) Fitness evaluation for each agent, 3) Update the fitness and acceleration for agents, 4) Calculate the new velocity and position, 6) If meet end of criterion, it is the best solution. The algorithm of GSA is stated below.

Gravitational Search Algorithm
Initialize the population
do
Evaluate the fitness for each agent
Calculate mass (m) and position (a) for each agent
Update the velocity and position
While (end)
User authenticated
End

Classification: A class to which it belongs based on a set of data is determined by Classification algorithms. One of the important areas of research is Classification of patterns and it has practical applications in a variety of fields, including pattern recognition, artificial intelligence and vision analysis. The typing pattern which is used to authenticate a user is by using keystroke dynamics a biometric based authentication. The user is genuine if the typing style during verification process matches the template stored in the database else it is imposter user. In authentication systems it is very important to validate whether the presented biometric matches the enrolled biometric of the same user. A variety of classification algorithms have been employed in this domain, including Statistical methods, Neural Network algorithms, Pattern recognition techniques and Fuzzy measure.

Adaptive Resonance Theory (ART) is an artificial neural network and is a pattern matching process that compares an external input with the internal memory of an active code [26, 2]. ART matching leads either to a resonant state, which persists long enough to permit learning, or to a parallel memory search. If the search ends at an established code, the memory representation may either remain the same or incorporate new information from matched portions of the current input. If the search ends at a new code, the memory representation learns the current input [4, 24]. This match- based learning process is the foundation of ART code stability. Match- based learning allows memories to change only when input from the external world is close enough to internal expectations, or when something completely new occurs. This feature makes ART systems well suited to problems that require online learning of large and evolving databases. Match- based learning is complementary to error- based learning, which responds to a mismatch by changing memories so as to reduce the difference between a target output and an actual output, rather than by searching for a better match. Error- based learning is

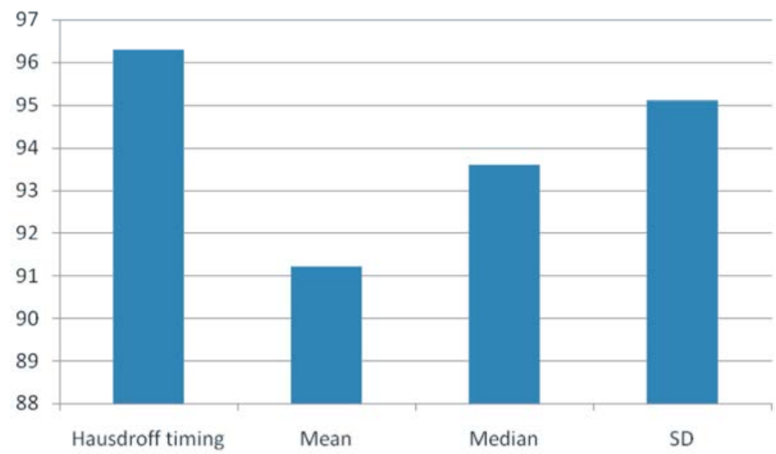

Fig. 3: Accuracy of preprocessing technique

Table 1: Latency feature data sheet of a user for the password "pass132" (6 samples)

S.No	pa	as	ss	s1	13	32	Mean	Median	S.D	Hausdorff timing
1	12.070	12.060	12.070	12.060	12.051	12.043	12.059	0.083	12.060	12.070
2	12.051	12.080	12.080	12.070	12.067	12.049	12.066	0.096	12.067	12.080
3	12.060	12.080	12.060	12.010	12.050	12.054	12.052	0.113	12.054	12.080
4	12.080	12.080	12.090	12.086	12.060	12.054	12.075	0.101	12.080	12.090
5	12.064	12.065	12.066	12.071	12.070	12.078	12.069	0.059	12.066	12.078
6	12.070	12.060	12.070	12.060	12.051	12.043	12.059	0.083	12.060	12.070

naturally suited to problems such as adaptive control and the learning of sensor- motor maps, which require ongoing adaptation to present statistics [19, 8].

Experimental Results: The proposed system is experimented with the dataset which represent the typing of 27 users on a single password. Duration, Latency, Digraph of every user were collected for the samples typed by every user and the values are stored in the template. The obtained sample is then used to calculate the mean, median, standard deviation and proposed Hausdorff timing for preprocessing. Table 1 shows the measured keystroke feature values of latency timing of a user for the password "pass132" of the samples and the corresponding mean, median, standard deviation and Hausdorff timing (distance) calculations.

The template for feature string which is used for further investigation is shown in Table 1. The initial populations for this template are 50 and during the test phase the maximum number of cycles was taken as 2000. The process is continued until the best fitted values are repeated. The performance of the algorithm was considered in terms of the best and average optimum values and the best solutions were reconsidered which becomes the input for the training of data using ART. The accuracy of preprocessing technique is shown in Figure 3.

In the experiment the input layer consists of four neurons, representing Hausdorff timing, mean, median and standard deviation obtained from the feature subset selection collection. The hidden layer consists of four neurons and the output layer is made up of one neuron. The learning rate was arbitrarily assigned to 0.6 and momentum term to 0.4. The appropriate parameter values are chosen based on trial and error performed during the experiment and on the convergence and goal performance result. This value is compared with target output of 0.1 and error value calculated. The adjusted weights between input to hidden and hidden to output are also calculated. The threshold value is obtained from maximum to minimum output within 28 iterations. Similarly twenty five weights are calculated and old weights of input to hidden layer are replaced after calculations. After training the user typing pattern, the threshold values for each trained user is assigned. Again the users were asked to verify by giving the password. The system was trained with 5 valid users and 5 invalid users. All invalid users were told valid passwords and try to get on to the system. After the verification of the password, the typing pattern is verified through the comparison of desired output with fixed threshold value. If the error value is less than 0.001 then the user is considered as valid user otherwise invalid user.

CONCLUSION

In this work the feature subset selection in keystroke dynamics based authentication used was gravitational search algorithm. Subsets of the feature were selected for the algorithm. The latency and the digraph Hausdorff timing provided the best performance by this algorithm.

REFERENCES

1. Grabowski Adam, 2003. On the Hausdorff distance between compact subsets. Formalized Mathematics, 11(2): 153-157.
2. Angadi, U.B. and Venkatesulu, 2010. Adaptive Resonance Theory-2 Neural network for protein classification.
3. Bleha, S.A. and M.S. Obaidat, 1991. Dimensionality reduction and feature extraction applications in identifying computer users. IEEE Transactions on Systems, Man and Cybernetics, 21(2): 452-456.
4. Dietrich Wienke and Lutgarde Buydens,1996. Adaptive resonance theory based neural network for supervised chemical pattern recognition (FuzzyARTMAP) Part 1: Theory and network properties. Chemometrics and Intelligent Laboratory Systems, pp: 151-164.
5. Lau Edmond, Xia Liu, Chen Xiao and Xiao Yu, 2004. Enhanced user authentication through keystroke biometrics. Tech. Bar. Massachusetts Institute of Technology, pp: 1-12.
6. Rashedi Esmat, Hossein Nezamabadi-pour and Saeid Saryazdi, 2009. GSA: A Gravitational Search Algorithm. Information Sciences, pp: 2232-2248.
7. Fadhli Wong Mohd Hasan Wong, Ainil Sufreena Mohd Supian and Ahmad Faris Ismail. 2001. Enhanced user authentication through typing biometrics with artificial neural networks and k-nearest neighbor algorithm. IEEE, pp: 911-915.
8. Carpenter Gail A. and Stephen Grossberg, 2002. Adaptive Resonance Theory. The Handbook of Brain Theory and Neural Networks, second edition, pp: 1-12.
9. Szeto, Grace P.Y., Leon M. Straker and Peter B. O'Sullivan, 2005. The effects of typing speed and force on motor control in symptomatic and asymptomatic office workers. International Journal of Industrial Ergonomics, pp: 779-795.
10. Hai- Rong, L.V., L.I.N. Zhong-Lin, Y.I.N. Wen-Jun and Jin Dong, 2008. Emotion recognition based on pressure sensor keyboards. IEEE, pp: 1089-1092.
11. Naji, Hamid Reza, Mina Sohrabi and Esmat Rashedi, 2012. A high-speed, performance-optimization algorithm based on a Gravitational approach. Computing in Science and Engineering, pp: 56-62.
12. Henrikson Jeff, 1999. Completeness and Total boundedness of the Hausdorff metric. Journal of Mathematics, (www-math.mit.edu/ phase2/ UJM/ vol1/ HAUSF.PDF), pp: 69-79.
13. Karnan, M, M. Akila and N. Krishnaraj, 2011. Biometric personal authentication using keystroke dynamics: A review. Applied Soft Computing, pp: 2565-1573.
14. Revett, Kenneth, Florian Gorunescu, Marina Gorunescu, Marius Ene, Sergio Tenreiro de Magalhaes and Henrique M.Dinis Santos, 2007. A machine learning aapproach to keystroke dynamics based user authentication. International Journal Electrical Security and Digital Forensics, 1(1): 55-70.
15. Chavent Marie, 2004. An Hausdorff distance between hyper- rectangles for clustering interval data. In: Classification, Clustering an Data Mining Applications, Proceedings of the IFCS'04, Springer, pp: 333-340.
16. Tang Min, Minkyoung Lee and Young J. Kin, 2009. Interactive Hausdorff distance computation for general polygonal models. ACM SIGGRAPH.
17. Bartlow Nick and Bojan Cukic, 2006. Evaluating the reliability of credential hardening through keystroke dynamics. IEEE.
18. Sabri, Norlina Mohd, Mazidah Puteh and Mohamad Rusop Mahmood, 2013. A review of Gravitational search algorithm, International Journal Advance Soft Computing Applications, 5(3): 1-39.
19. Pham, D.T. and A.B. Chan, 2011. Unsupervised adaptive resonance theory neural networks for control chart pattern recognition, Proceedings Instrumentation Mechanical Engineering, 215: 59-67.
20. Pilsung Kang, Park Sunghoon, Sungzoon Cho, Seong-seob Hwang and Hyoung-joo Lee, 2006. The effectiveness of Artificial rhythms and cues in keystroke dynamics based user authentication. Springer- Verlag, pp: 161-162.
21. Raphael Straub. 2007. Exact computation of the Hausdorff distance between triangular meshes. EG Short Papers, pp: 17-20.
22. Rozali, Sahazati Md, Mohd Fua ad Rahmat and Abdul Rashid Husain, 2014. Performance comparison of Particle Swarm Optimization and Gravitational Search Algorithm to the designed of controller for nonlinear system. Journal of Applied Mathematics, pp: 1-9.

23. Mandujano Salvador and Rogelio Soto, 2004. Deterring password sharing: user authentication via Fuzzy c-Means clustering applied to keystroke biometric data. IEEE.
24. Grossberg Stephen, 2013. Adaptive Resonance Theory: How a brain learns to consciously attend, learn and recognize a changing world. Neural Networks, pp: 1-47.
25. Chengyi Tong, 2014. Gravitational search algorithm based on simulated annealing. Journal of Convergence Information Technology (JCIT), 9(2): 231-237.
26. Wienke, D., Broek Van den, L. Buuydens, T. Huth-Fehre, R. Feldhoff, T. Kantimm and K. Cammann, 1996. Adaptive resonance theory based neural network for supervised chemical pattern recognition (FuzzyARTMAP) Part 2: Classification of post-consumer plastics by remote NIR spectroscopy using InGaAs diode array. Chemometrics and Intelligent Laboratory Systems, pp: 165-176.
27. Rucklidge William, J., 1997. Efficiently locating objects using the Hausdorff distance. International Journal of Computer Vision, pp: 251-270.
28. Jadidi, Zahra, Vallaipuram Muthukkumarasamy, Elankayer Sithirasenan and Mansour Sheikhan, 2013. Flow- based anomaly detection using neural network optimized with GSA algorithm. IEEE, pp: 76-81.